Article

Prognostic Gene Expression-Based Signature in Clear-Cell Renal Cell Carcinoma

Fiorella L. Roldán [1,2,†], Laura Izquierdo [1,2,†], Mercedes Ingelmo-Torres [1,2], Juan José Lozano [3], Raquel Carrasco [1,2], Alexandra Cuñado [1], Oscar Reig [4], Lourdes Mengual [1,2,5,*,‡] and Antonio Alcaraz [1,2,‡]

1   Laboratori i Servei d'Urologia, Hospital Clínic de Barcelona, 08036 Barcelona, Spain;
    flroldan@clinic.cat (F.L.R.); lizquier@clinic.cat (L.I.); ingelmo@clinic.cat (M.I.-T.); racarrasco@clinic.cat (R.C.);
    scunado@clinic.cat (A.C.); aalcaraz@clinic.cat (A.A.)
2   Genètica i Tumors Urològics, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS),
    08036 Barcelona, Spain
3   Plataforma de Bioinformàtica, Centro de Investigación Biomédica en Red Enfermedades Hepáticas y
    Digestivas (CIBERehd), Hospital Clínic, 08036 Barcelona, Spain; juanjo.lozano@ciberehd.org
4   Servei d'Oncologia Mèdica, Hospital Clínic de Barcelona, 08036 Barcelona, Spain; oreig@clinic.cat
5   Departament de Biomedicina, Facultat de Medicina i Ciències de la Salut, Universitat de Barcelona (UB),
    08036 Barcelona, Spain
*   Correspondence: lmengual@clinic.cat; Tel.: +34-93-227-54-00 (ext. 4820)
†   These authors contributed equally to this work.
‡   These authors contributed equally to this work.

Simple Summary: In this study, we identified molecular markers for disease progression from ccRCC tissue samples. Using the selected biomarkers and clinical data from the TCGA cohort, we developed a gene expression-based signature which enhances the prognostic prediction of clinicopathological variables and could help to provide personalized disease management.

Abstract: The inaccuracy of the current prognostic algorithms and the potential changes in the therapeutic management of localized ccRCC demands the development of an improved prognostic model for these patients. To this end, we analyzed whole-transcriptome profiling of 26 tissue samples from progressive and non-progressive ccRCCs using Illumina Hi-seq 4000. Differentially expressed genes (DEG) were intersected with the RNA-sequencing data from the TCGA. The overlapping genes were used for further analysis. A total of 132 genes were found to be prognosis-related genes. LASSO regression enabled the development of the best prognostic six-gene panel. Cox regression analyses were performed to identify independent clinical prognostic parameters to construct a combined nomogram which includes the expression of *CERCAM*, *MIA2*, *HS6ST2*, *ONECUT2*, *SOX12*, *TMEM132A*, pT stage, tumor size and ISUP grade. A risk score generated using this model effectively stratified patients at higher risk of disease progression (HR 10.79; $p < 0.001$) and cancer-specific death (HR 19.27; $p < 0.001$). It correlated with the clinicopathological variables, enabling us to discriminate a subset of patients at higher risk of progression within the Stage, Size, Grade and Necrosis score (SSIGN) risk groups, pT and ISUP grade. In summary, a gene expression-based prognostic signature was successfully developed providing a more precise assessment of the individual risk of progression.

Keywords: gene expression; clear-cell renal cell carcinoma; disease progression; prognostic factors; biomarkers; RNA sequencing

1. Introduction

Renal cell carcinoma (RCC) ranks third among the urological cancers with the highest incidence. Over 431,000 new cases and more than 170,000 RCC-related deaths were reported worldwide last year [1]. Clear-cell RCC (ccRCC) is the most common histological subtype and has the worst prognosis among all RCCs [2]. Currently, most of the newly diagnosed ccRCC cases are organ-confined tumors; however, after curative treatment, up to 30–40%

will develop tumor metastases. Unfortunately, metastatic patients have a very poor five-year survival rate, varying between 0–20% [3,4].

According to the European Urological Guidelines, the standard management for all localized ccRCCs receiving surgery is limited to radiographic surveillance. No adjuvant treatment is approved for patients with a higher risk of progression [2]. Moreover, all surveillance recommendations are only based on clinical parameters, even though these have been proven insufficient to accurately predict disease progression, either to select ccRCC patients for adjuvant treatments or guide disease management [5,6].

Gene expression profiling has been used extensively in cancer research and has led to the discovery of new molecular prognostic markers and potential therapeutic targets. Several genetic models have been proposed in ccRCC [7–10]; however, none of those classifiers have been widely accepted nor implemented in routine clinical practice. Biomarker research for ccRCC still faces multiple challenges, mainly due to tumor heterogeneity and lack of validation studies. In addition, the use of high-throughput assays and the identification of a significant number of markers in a relatively small number of patients increase the complexity of data analysis [3]. The currently validated gene signatures comprise a large number of biomarkers, hindering their applicability and reproducibility [8,9]. Therefore, in this study we sought to develop a novel and high-performing gene expression-based signature using data generated from our cohort and The Cancer Genome Atlas (TCGA) cohort, to provide a more accurate assessment of the individual risk of progression for patients with localized ccRCC.

2. Materials and Methods

2.1. Patients, Datasets Sources and Study Design

This study was split into a three-stage approach: an initial molecular profiling, a selection and verification of prognosis-related genes and a signature development phase (Figure 1). The initial molecular profiling phase included a total of 26 localized ccRCCs who underwent partial or radical nephrectomy between 2001 and 2010 in our center (Hospital Clinic of Barcelona, Barcelona, Spain). These 26 cases consisted of 13 progressive and 13 non-progressive patients and met the following criteria: no neoadjuvant or adjuvant treatment, no prior or concomitant malignancies or a history of inherited von Hippel-Lindau disease, all patients had thoracoabdominal CT scan staging within two month before surgery to ensure organ-confined disease. Tumors were considered progressive when local relapse or distant metastasis developed during follow-up. All patients were followed up postoperatively according to the European Urology guidelines. Any progressive patient within two months of surgery was excluded from the study. Non-progressive patients had a minimum follow-up of 10 years to ensure their status as appropriate controls. Tissue samples were obtained under institutional review board-approved protocols (HBC/2016/0333).

The selection of prognosis-related genes and signature development phases was carried out using The Cancer Genome Atlas (TCGA) dataset. Level 3 RNAseq expression data and the corresponding clinical data from TCGA ccRCC samples were obtained from the portal (https://firebrowse.org (accessed on 8 October 2021) [11] (Supplementary Material Table S1). Survival data were obtained from portal (https://www.sciencedirect.com/science/article/pii/S0092867418302290?via%3Dihub#app2 (accessed on 8 October 2021) [12]. After selecting samples matching our selection criteria and excluding patients without survival status or missing clinical data, a total of 356 ccRCC samples from TCGA, 68 progressive and 288 non-progressive, were selected and the gene expression of 20,532 genes was downloaded.
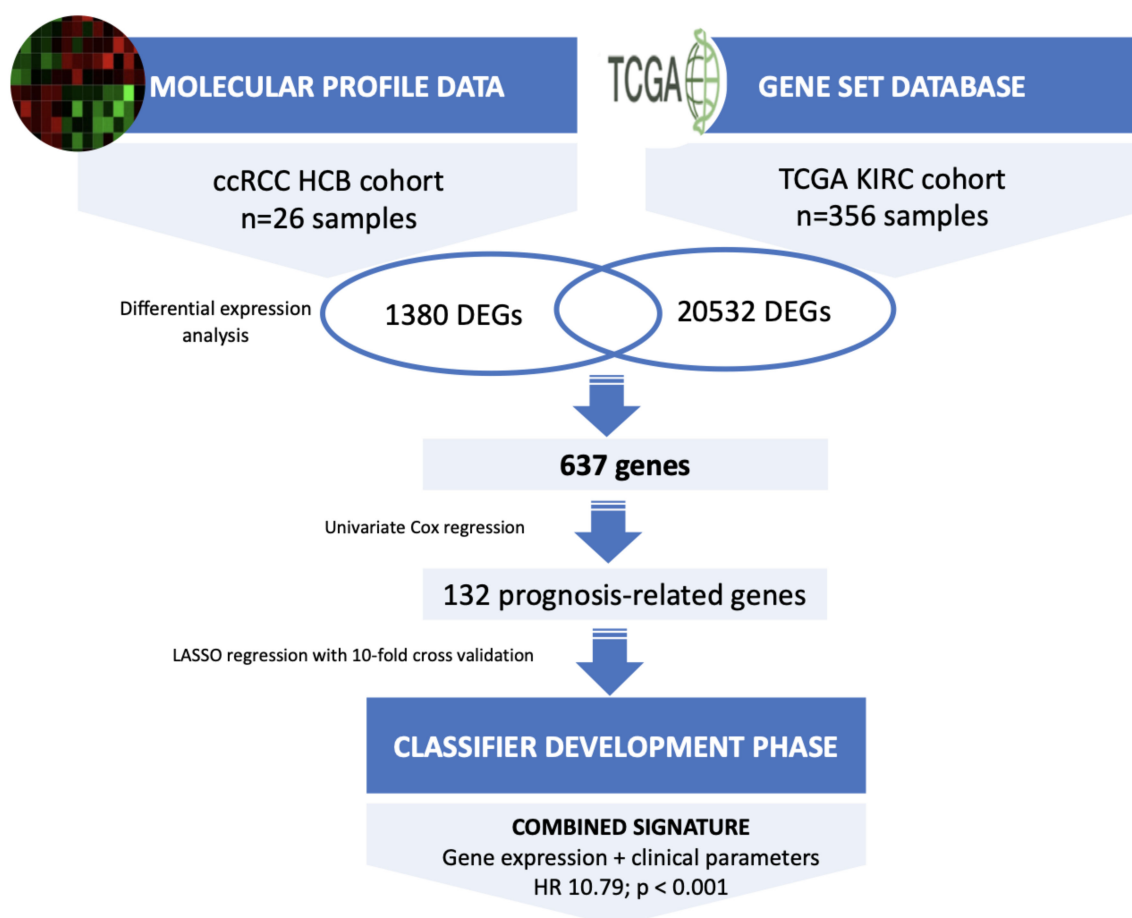
Figure 1. Flowchart of the whole study. Abbreviations: ccRCC, clear-cell renal cell carcinoma; DEGs, differentially expressed genes.

2.2. Tissue Specimens and RNA Isolation

Formalin-fixed paraffin-embedded (FFPE) tissue blocks were reviewed. The tumor area was macro-dissected from slides (total thickness 80 μm) and RNA was isolated from FFPE specimens using the kit RecoverAll™ Total Nucleic Acid Isolation for FFPE (Ambion, Inc. Austin, TX, USA), following manufacturers' instructions. RNA was quantified by spectrophotometric analysis at 260 nm (NanoDrop Technologies, Wilmington, DE, USA) and RNA integrity was assessed using Agilent 2100 Bioanalyzer System.

2.3. Molecular Profiling by RNA Sequencing

Library preparation and sequencing method: Following rRNA removal (Ribo-Zero® rRNA Removal Kit, Illumina), RNA from 26 selected ccRCC samples was processed for library preparation using the TruSeq® RNA Access Library Preparation Kit (Illumina, San Diego, CA, USA) that allows generating libraries starting from degraded RNA. Briefly, cDNA strands were synthetized from input RNA in order to be adaptor-tagged, labeled and amplified. cDNA was then pooled and enriched by a double step of probes hybridization. The enriched targets were captured by streptavidin labeled beads, cleaned up and amplified to obtain the final multiplexed libraries. The libraries were then sequenced on an Illumina HiSeq® 4000 platform (Illumina®).

Read alignment and differential gene expression analysis: Paired-end RNA-Seq FASTQ files were trimmed from a 3′ end to a fixed length based on the Phred quality score (trimmed if score fell below 20, with a minimum read length of 25) [13]. Trimmed RNA-seq reads were aligned to the GRCh38 reference genome with STAR [14] and gene counts were determined using quantMode GeneCounts. Trimmed reads were then aligned using STAR. We used

limma-voom transformation and cyclic-loess to normalize the non-biological variability. An assessment of differential expression between groups was evaluated using moderated t-statistics [15].

Significant DEGs between progressive and non-progressive patients were identified based on an adjusted *p*-value of <0.05 and a fold change (FC) $\geq \pm 2$. The heatmap and statistical analyses were performed using the R statistical package (v3.3.2). Gene set enrichment analysis (GSEA) was performed using GSEA2-2.2.0 software for testing specific gene sets based on Gene Ontology (GO) Biological Processes [16]. The "EnrichmentMap" plug-in of Cytoscape was used to create an enrichment map of the GSEA results, depicting the overlap among pathways, with similar biological processes grouped together as subnetworks [17–19]. A conservative overlap coefficient (0.5) was used to build the enrichment map. The "AutoAnnotate" plug-in identified clusters in an automated manner, visually annotating them with a summary label [20]. RNAseq files and clinical information were deposited in the Gene Expression Omnibus (GEO) with accession number GSE175648.

2.4. Selection and Verification of Prognosis-Related Genes

The DEGs identified in the previous phase were intersected with the TCGA gene expression dataset and overlapping genes were used for further analysis. The raw counts of RNA-sequencing data from the TCGA cohort were normalized using log2-based transformation. This normalized expression was used to build a multigene signature panel. Firstly, we performed univariate Cox regression analysis (considering 637 genes and 3 clinical variables) to identify the potential prognosis-related variables. Then, LASSO (Least Absolute Shrinkage and Selection Operator) regression was applied to construct the gene-based signature for predicting tumor progression in ccRCC using the cvr.glmnet function from ipflasso R package using ten-fold cross-validation and repeated it ten successive runs to increase reliability and robustness [21]. In the machine learning procedure, we fixed the three clinical variables. For all statistical analysis, a *p*-value < 0.05 was considered significant. All statistical analyses were performed using SPSS 19.0 (Statistical Product and Service Solutions; IBM Corporation, Armonk, NY, USA) and R version 3.4.2 (R Foundation for Statistical Computing, Vienna, Austria).

2.5. Development of a Gene Expression-Based Signature

Based on the expression of each gene discovered and the three clinical variables, each patient's risk score (RS) was calculated according to the risk score model. The risk score model was then used to evaluate the ccRCC prognosis according to the general form RS = exp $\Sigma \beta_i x_i s$, where $i = 1$, k index variables, β_i represents the coefficient for each variable estimated from the Cox regression model, and $x_i s$ the corresponding value for each variable in a given patient. RS was subjected to a Receiver Operating Characteristics (ROC) curve analysis to choose the most appropriate threshold for predicting tumor progression. Thereafter, Kaplan–Meier curves were generated using the selected cut-off point and compared according to the log-rank test.

The endpoints were disease progression, defined as any local relapse or distant metastasis demonstrated by radiological imaging, and cancer-specific survival. We investigated the role of the gene panel alone, clinicopathological variables alone, and a combined model including gene expression and clinicopathological variables as potential predictors of disease progression and cancer-specific survival.

2.6. Pathway Enrichment Analysis

Ingenuity Pathway Analysis (IPA) software was used to identify interactions and networks between the prognostic markers included in our gene signature, possible altered canonical pathways, regulators, diseases and functions based on direct/indirect and experimental targets.

3. Results

3.1. Clinical Features of the Cohort

The clinicopathological characteristics of patients divided by study phase are summarized in Table 1. The median (range) follow-up of the cohort was 45.6 (2.1–135.8) months. During the follow-up period, a total of 68 patients (19.1%) developed tumor progression and a total of 36 patients died of ccRCC. The median time to tumor progression and cancer-related death was 19.9 (2.1–125.5) and 48.5 (2.5–151.2) months, respectively.

Table 1. Demographic and pathological characteristics of enrolled patients.

KERRYPNX	Discovery Phase Hospital Clinic Barcelona (*n* = 26)	Validation Phase TCGA Cohort (*n* = 356)
Gender		
Male	18 (69.2)	231 (64.9)
Female	8 (30.8)	125 (35.1)
Age at diagnosis (year)	59 (34–81)	60 (29–90)
Pathological tumor size (cm)	5.5 (1.9–17.5)	5.1 (1.0–25)
ISUP		
ISUP 1	3 (11.5)	4 (1.1)
ISUP 2	12 (46.2)	173 (48.6)
ISUP 3	6 (23.1)	145 (40.7)
ISUP 4	5 (19.2)	34 (9.6)
Tumor stage		
pT1	15 (57.7)	211 (59.3)
pT2	5 (19.2)	41 (11.5)
pT3	5 (19.2)	102 (28.7)
pT4	1 (3.8)	2 (0.6)
N stage		
N0/x	24 (92.3)	346 (97.2)
N1	2 (7.7)	10 (2.8)
Necrosis	10 (38.5)	144 (40.4)
SSIGN score *		
Low risk	12 (46.2)	143 (40.2)
Intermediate risk	8 (30.7)	141 (39.6)
High risk	6 (23.1)	72 (20.2)

* Stage, Size, Grade and Necrosis (SSIGN) score [22].

3.2. Molecular Profiling of ccRCC Samples

Overall, we identified 1380 transcripts that were differentially expressed (*p* < 0.05) between progressive and non-progressive ccRCC samples. Of these, 639 were protein-coding genes; 217 were downregulated and 422 upregulated in progressive compared with non-progressive cases. A heat map based on the most DEGs between the two groups of ccRCC patients is shown in Figure 2A. Gene set enrichment analysis (GSEA) identified several enriched biological processes, such as the dependent toll-like receptor signaling pathway, metabolic process and immune response regulating cell surface receptor signaling pathway (Figure 2B). The full list of GO biological processes is available in Supplementary Materials (Table S2). To aid interpretation of these enriched pathways, we used enrichment maps to create a network-based representation of our results. The most prominent cluster of significantly enriched pathways recapitulated changes in the Catabolism Biological Process and Toll Signaling Pathway (Figure 2C).
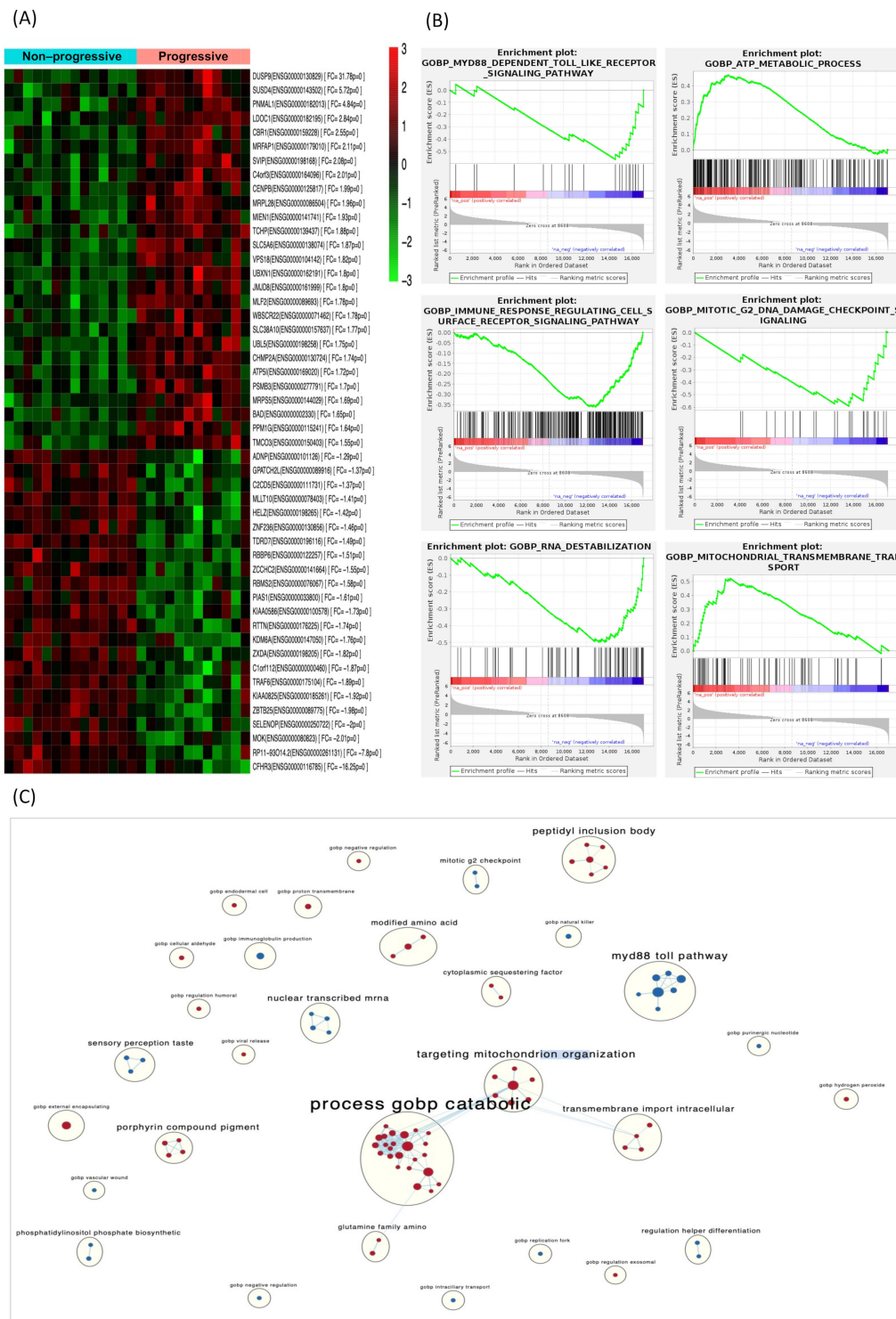
Figure 2. DEGs in the discovery phase and gene-set enrichment analysis. (**A**) Heat map displaying the 50 most DEGs between progressive and non-progressive localized ccRCC patients. Red pixels correspond to upregulated genes, whereas green pixels indicate downregulated genes. (**B**) GSEA shows positive correlation of DEGs in biological processes involved in tumor progression. (**C**) Enrichment map where nodes represent gene sets (pathways) and edges (blue lines) denote overlapping genes between 2 pathways. Node size denotes gene set size. Predicted pathways are grouped as circles, where shades in red correspond to up-regulated gene-sets and shades in light blue correspond to down-regulated gene-sets. Highly redundant gene sets are grouped together as clusters. Abbreviations: DEGs, differentially expressed genes. GSEA, gene set enrichment analysis.

3.3. Identification of Prognosis-Related Genes in an External Data Set

To validate the 639 genes identified as DEGs in the previous study phase, these genes were intersected with the 20,532 genes from the TCGA cohort (Figure 1). As a result, we obtained 637 overlapping genes. Of those, univariate Cox regression analysis identified 132 prognosis-related DEGs and three clinicopathological variables (pT stage, tumor size and ISUP grade).

LASSO regression analysis was used to select the best combination of genes significantly associated with disease progression and to build a six-gene signature. The expressions of five of these genes: *CERCAM*, *HS6ST2*, *ONECUT2*, *SOX12* and *TMEM132A* were upregulated, while *MIA2* expression was downregulated in progressive compared with non-progressive cases. According to Ingenuity Pathway Analysis (IPA), these six validated genes were enriched in cancer, organismal injury, abnormalities, cell-to-cell signaling and interactions, cell-mediated immune response, cellular development, cellular growth and proliferation, carbohydrate metabolism, and angiogenesis, among others. Significant IPA canonical pathways are depicted in Supplementary Material Table S3. The network generated shows that there were no direct interactions between the six prognostic genes (Figure S1).

Gene expression values for each selected gene were used for Cox regression analysis. High expression of *CERCAM*, *HS6ST2*, *ONECUT2*, *SOX12* and *TMEM132A* and low expression of *MIA2* related to poor outcomes for progression-free survival (Figure S2) and cancer-specific survival (Table 2). Moreover, the clinicopathological variables pT stage, tumor size and ISUP were also found as to be prognostic factors for both survival endpoints.

Table 2. Univariate Cox regression analysis of statistically significant genetic and clinical variables in the validation set (TCGA cohort).

|	Progression-Free Survival			Cancer-Specific Survival		
	p	95% CI	HR	*p*	95% CI	HR
CERCAM	<0.001	1.387–3.807	2.298	<0.001	1.036–1.075	1.055
HS6ST2	<0.001	1.164–3.106	1.902	0.034	1.043–2.866	1.729
MIA2	<0.001	0.222–0.632	0.375	<0.001	0.825–0.935	0.878
ONECUT2	0.015	1.111–2.952	1.811	<0.001	2.443–5.942	3.810
SOX12	0.001	1.354–3.748	2.252	<0.001	1.177–1.488	1.323
TMEM132A	<0.001	1.526–4.288	2.558	<0.001	1.070–1.156	1.112
pT Stage	<0.001	1.775–3.024	2.317	<0.001	2.547–9.940	5.032
Tumor size	<0.001	1.154–1.273	1.212	<0.001	1.125–1.271	1.195
ISUP	<0.001	1.568–3.158	2.225	0.001	1.628–7.845	3.574

3.4. Development of a Prognostic Signature

The risk score (RS) for disease progression was calculated for each patient according to a mathematical algorithm containing the six-gene expression values; pT stage, tumor size and ISUP grade (Supplementary Material Table S4). An ROC analysis of this combined gene expression–clinicopathological model was performed and allowed the selection of a threshold of 0.789 (sensitivity 90% and specificity 60%) and 0.799 (sensitivity 94% and specificity 55%) to categorize patients into high- and low-risk groups for tumor progression and cancer-related death, respectively. The Kaplan–Meier curve of the combined generated gene expression-based model was able to discriminate two groups with significantly different probabilities of tumor progression (hazard ratio (HR) 10.79; 95%, $p < 0.001$) and cancer-specific survival (HR 19.27; 95%, $p < 0.001$) (Figure 3). Notably, the performance of the combined gene expression-based model (Area under the Curve [AUC] 0.824) was higher than that of clinicopathological variables alone (AUC 0.766) or gene expression data alone (AUC 0.753) (Figure S3).
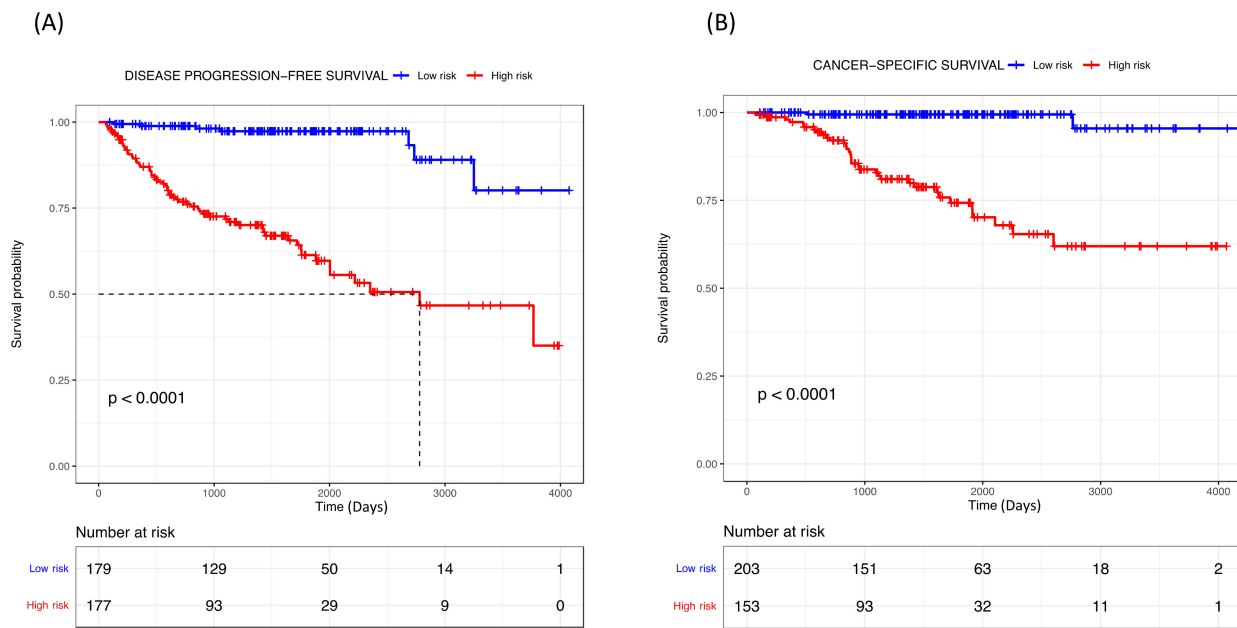
Figure 3. Kaplan–Meier curves of the combined gene expression-based model for (**A**) disease progression-free survival and (**B**) cancer-specific survival for TCGA cohort.

3.5. Correlation Analysis of the RS with Clinical Characteristics for Disease Progression

Given the clinical significance of the RS in ccRCC, we sought to investigate the potential correlation between RS and clinical features. The Mann–Whitney test revealed that higher RSs correlated with a higher risk group within the SSIGN model, higher pT stage and higher ISUP grade (Figure 4). Furthermore, the Kaplan–Meier curve indicated that our established RS was capable of identifying ccRCC patients at the highest risk of progression within the groups stratified by SSIGN, pT stage and ISUP grade (all $p < 0.05$; Figure S4).
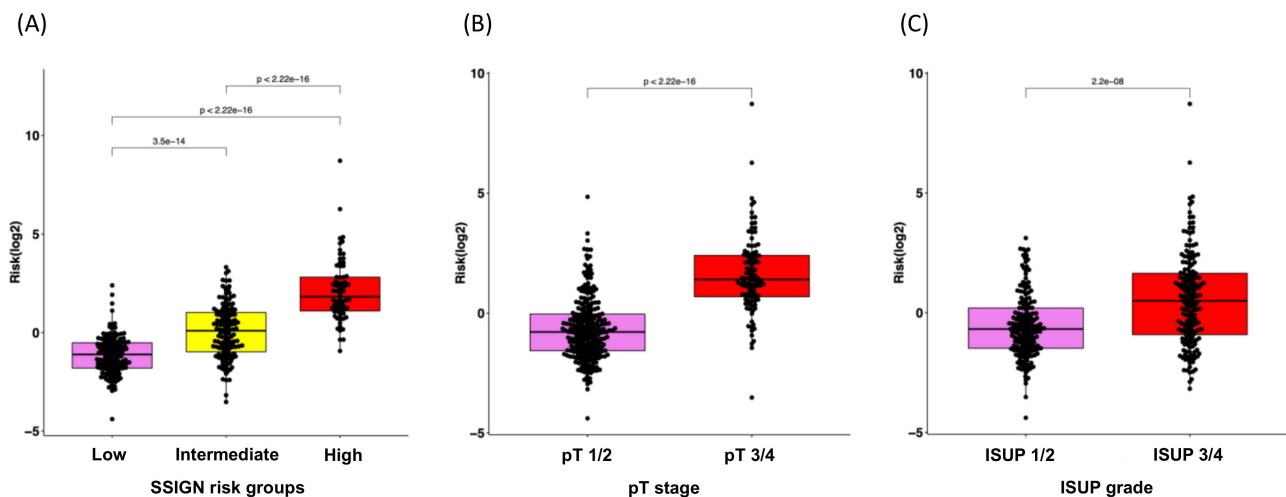
Figure 4. Box plots for the correlation analysis of RS with clinical characteristics for disease progression. (**A**) SSIGN risk groups, (**B**) pT stage and (**C**) ISUP grade.

4. Discussion

Currently, clinicopathological variables are the most valuable tool for predicting disease outcomes in ccRCC. However, due to the highly variable behavior of ccRCC, the prediction of tumor progression is still an important clinical challenge. For decades, surgery

has remained the only treatment approved for localized ccRCC [2]. At present, disease management of these patients at high risk of progression is changing and new adjuvant treatments are being considered, opening a door to more personalized medicine in localized renal carcinoma [23,24].

Molecular profiling helps our understanding of the molecular mechanism underlying ccRCC and affords great potential to identify new biomarkers of clinical utility [25,26]. However, intratumor heterogeneity and the methodology for sample processing, readout and expression normalization have been a strong challenge in the development of a robust gene signature. Next-generation sequencing is the most advanced technique for gene expression profiling [27]; here, we used this technology to analyze the entire transcriptome profiling of tissue samples from progressive and non-progressive ccRCC patients. We then established a gene signature for predicting disease progression based on the gene expression and clinical data obtained from the TCGA cohort. We improved the gene selection and accuracy of the model by using LASSO regression analysis; this allowed us to include all DEGs found in the training set and avoid the preselection and validation of only a subset of these DEGs [28].

This study demonstrated that our gene expression-based signature was able to identify localized ccRCC patients with high and low risk of disease progression in the whole cohort and within the SSIGN risk groups. It properly correlated with clinical parameters and was proven to enhance the predictive value of the current clinicopathological variables. Furthermore, it was also predictive of cancer-specific survival. Therefore, our signature may constitute an important step forward in treatment decisions for ccRCC patients.

Remarkably, the developed gene-based panel demonstrated a greater value for prognostic prediction (HR 10.79, $p < 0.001$; AUC = 0.824) compared with similar, previously described models. Dai et al. proposed a four-gene signature with an HR of 3.1, $p < 0.001$, whereas Zhao et al. described a 15-gene model with an AUC of 0.737 [29]. Unfortunately, the different designs and methodologies of several other studies thwart any performance comparisons with their proposed genetic models. Thus, Brook et al. [9] assessed the performance of ClearCode34, which classifies ccRCC into ccA and ccB subtypes, ccB presented tumor relapse more frequently (HR 2.1; $p = 0.001$), whereas Rini et al. [8] generated a 16-gene signature associated with tumor recurrence with an HR per 25-unit increase in score of 3.37 ($p < 0.001$). Likewise, other authors have built molecular signatures aiming to predict overall survival (OS), thus making the classifiers' performance non-comparable [30–32].

Biologically, the genes from our panel are unrelated to each other and many of them have been shown to have either prognostic or biologic relevance in tumor metastasis development. According to previous reports, some of our selected genes were consistent with previously discovered biomarkers; therefore, we have further validated their value for ccRCC progression. Briefly, *CERCAM* (cerebral endothelial cell adhesion molecule) is an adhesion molecule found to be an unfavorable prognostic marker in several tumors [33–35]. Its overexpression promotes cell viability, proliferation and invasion, it is involved in the PI3K/AKT pathway [36] and is part of an immune prognostic signature for colon and rectal cancer [35]. *HS6ST2* (heparan sulfate D-glucosaminyl 6-*O*-sulfotransferase-2) is a glycolysis-related gene and has also been associated with poor disease outcomes in numerous malignancies [10,37]. Interestingly, our group previously validated this gene as an independent prognosis biomarker in intermediate/high-risk ccRCC and found it to be associated with angiogenesis, epithelia–mesenchymal transition (EMT), and indirectly related to the PD-1, PDL-1 cancer immunotherapy pathway [37–39]. *MIA2* (melanoma inhibitory activity 2) has been found in several malignancies and can act as a tumor suppressor [40] or as a proto-oncogene depending on the receptor-related signaling differences [41]. We found *MIA2* to be downregulated in progressive ccRCC. This is congruent with the human protein atlas findings, where high expression was a favorable prognostic factor in renal cancer [33,42]. *ONECUT2* (One cut domain family member 2) is a transcription factor able to activate oncogenic pathways and lineage-specific genes; hence, it is involved in EMT, angiogenesis, neural differentiation, proliferation, extracellular matrix organization, cell

locomotion and migration, among others. Overexpression of *ONECUT2* has been described in several tumors and is related to poor prognosis [43–46]. *SOX12* (Sex-determining region Y-box12) is a transcription factor, its upregulation promotes tumor progression and it is involved in EMT, apoptosis and cell proliferation [47–49]. It functions as an oncogene-regulating Wnt/B-catenin signaling to promote the growth of multiple myeloma cells [50]. As for *TMEM132A*, few studies were found in the literature, so further investigations are required to establish its role in tumor development.

Our study has multiple strengths. The first advantage of our gene expression-based signature is that it contains a low number of genes, making its clinical application easier. Our model did match genes from previous models and some of them exceeded the mere field of ccRCC, highlighting the prognostic power of the selected genes. The fact that they are involved in different carcinogenic mechanisms confers an advantage to our signature compared with others that only target one single pathway [10,51]. The high-throughput technology used to analyze the samples and the statistical methodology makes our gene model a reliable tool for predicting disease progression in ccRCC and adds important prognostic information to the clinicopathological parameters. However, we acknowledge that this study has several limitations. First, the retrospective design and the relatively small sample might have influenced our findings. Second, the definition of CSS in the TCGA cohort should be taken with caution since it was estimated [12]. Finally, despite the good performance of our six-gene model, further validations in larger cohorts are required.

5. Conclusions

A gene expression-based prognostic signature to predict disease progression in ccRCC was successfully developed; it could discriminate two groups with different probabilities of tumor recurrence. In addition, our model was also useful in predicting cancer-specific survival. The combination of genetic and clinical information enhanced the current risk stratification of the localized ccRCC patients. Refining prognostic algorithms could help to improve the disease management and follow-up of ccRCC patients.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14153754/s1, Figure S1. Network of gene-gene interactions for the six genes from the gene expression-based signature. Figure S2. Kaplan–Meier curves of each gene of the model evaluating progression free survival (data from the TCGA cohort). Figure S3. ROC curves showing the performance of the gene expression-based signature. The figure depicts that the combined gene expression-based model outperforms clinicopathological variables alone or gene expression data alone. Figure S4. Kaplan–Meier curves of the combined gene expression-based model to evaluate progression-free survival within (A) SSIGN risk groups (low, intermediate and high), (B) pT stage groups (pT1/2 and pT3/4) and (C) ISUP grade groups (ISUP 1/2 and ISUP 3/4) (data from the TCGA cohort). Table S1. Clinical database of TCGA cohort used in this study. Table S2. List of GO biological processes from 1380 differentially expressed transcripts between progressive and non-progressive ccRCC samples. Table S3. Significant IPA canonical pathways from the six genes of the signature. Table S4. Risk Score for disease progression.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Hospital clinic of Barcelona (protocol code HBC/2016/0333 and date of approval was 24 January 2019).

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Ljungberg, B.; Albiges, L.; Bensalah, K.; Bex, A.; Giles, R.H.; Hora, M.; Kuczyk, M.A.; Lam, T.; Marconi, L.; Canfield, S.; et al. EAU Guidelines. Edn. presented at the EAU Annual Congress Amsterdam 2020. *Eur. Urol.* **2020**, *67*, 913–924. [CrossRef] [PubMed]
3. Klatte, T.; Rossi, S.H.; Stewart, G.D. Prognostic factors and prognostic models for renal cell carcinoma: A literature review. *World J. Urol.* **2018**, *36*, 1943–1952. [CrossRef] [PubMed]
4. Padala, S.A.; Barsouk, A.; Thandra, K.C.; Saginala, K.; Mohammed, A.; Vakiti, A.; Rawla, P.; Barsouk, A. Epidemiology of Renal Cell Carcinoma. *World J. Oncol.* **2020**, *11*, 79. [CrossRef]
5. Motzer, R.J.; Haas, N.B.; Donskov, F.; Gross-Goupil, M.; Varlamov, S.; Kopyltsov, E.; Lee, J.L.; Melichar, B.; Rini, B.I.; Choueiri, T.K.; et al. Randomized phase III trial of adjuvant pazopanib versus placebo after nephrectomy in patients with localized or locally advanced renal cell carcinoma. *J. Clin. Oncol.* **2017**, *35*, 3916–3923. [CrossRef]
6. Haas, N.B.; Manola, J.; Uzzo, R.G.; Flaherty, K.T.; Wood, C.G.; Kane, C.; Jewett, M.; Dutcher, J.P.; Atkins, M.B.; Pins, M.; et al. Adjuvant sunitinib or sorafenib for high-risk, non-metastatic renal-cell carcinoma (ECOG-ACRIN E2805): A double-blind, placebo-controlled, randomised, phase 3 trial. *Lancet* **2016**, *387*, 2008–2016. [CrossRef]
7. Zhao, H.; Cao, Y.; Wang, Y.; Zhang, L.; Chen, C.; Wang, Y.; Lu, X.; Liu, S.; Yan, F. Dynamic prognostic model for kidney renal clear cell carcinoma (KIRC) patients by combining clinical and genetic information. *Sci. Rep.* **2018**, *8*, 17613. [CrossRef]
8. Rini, B.; Goddard, A.; Knezevic, D.; Maddala, T.; Zhou, M.; Aydin, H.; Campbell, S.; Elson, P.; Koscielny, S.; Lopatin, M.; et al. A 16-gene assay to predict recurrence after surgery in localised renal cell carcinoma: Development and validation studies. *Lancet Oncol.* **2015**, *16*, 676–685. [CrossRef]
9. Brooks, S.A.; Brannon, A.R.; Parker, J.S.; Fisher, J.C.; Sen, O.; Kattan, M.W.; Hakimi, A.A.; Hsieh, J.J.; Choueiri, T.K.; Tamboli, P.; et al. ClearCode34: A prognostic risk predictor for localized clear cell renal cell carcinoma. *Eur. Urol.* **2014**, *66*, 77–84. [CrossRef]
10. Xing, Q.; Zeng, T.; Liu, S.; Cheng, H.; Ma, L.; Wang, Y. A novel 10 glycolysis-related genes signature could predict overall survival for clear cell renal cell carcinoma. *BMC Cancer* **2021**, *21*, 381. [CrossRef]
11. FireBrowse. Available online: http://firebrowse.org/ (accessed on 8 October 2021).
12. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A.V.; et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **2018**, *173*, 400–416.e11. [CrossRef]
13. Babraham Bioinformatics—Trim Galore! Available online: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ (accessed on 15 May 2019).
14. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
15. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [CrossRef]
16. Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef]
17. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef]
18. Merico, D.; Isserlin, R.; Stueker, O.; Emili, A.; Bader, G.D. Enrichment map: A network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE* **2010**, *5*, e13984. [CrossRef]
19. Reimand, J.; Isserlin, R.; Voisin, V.; Kucera, M.; Tannus-Lopes, C.; Rostamianfar, A.; Wadi, L.; Meyer, M.; Wong, J.; Xu, C.; et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat. Protoc.* **2019**, *14*, 482–517. [CrossRef]
20. Kucera, M.; Isserlin, R.; Arkhangorodsky, A.; Bader, G.D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Research* **2016**, *5*, 1717. [CrossRef]

21. Boulesteix, A.L.; De Bin, R.; Jiang, X.; Fuchs, M. IPF-LASSO: Integrative L1-Penalized Regression with Penalty Factors for Prediction Based on Multi-Omics Data. *Comput. Math. Methods Med.* **2017**, *2017*, 7691937. [CrossRef]

22. Frank, I.; Blute, M.L.; Cheville, J.C.; Lohse, C.M.; Weaver, A.L.; Zincke, H. An outcome prediction model for patients with clear cell renal cell carcinoma treated with radical nephrectomy based on tumor stage, size, grade and necrosis: The SSIGN score. *J. Urol.* **2002**, *168*, 2395–2400. [CrossRef]

23. Choueiri, T.K.; Tomczak, P.; Park, S.H.; Venugopal, B.; Ferguson, T.; Chang, Y.-H.; Hajek, J.; Symeonides, S.N.; Lee, J.L.; Sarwar, N.; et al. Adjuvant Pembrolizumab after Nephrectomy in Renal-Cell Carcinoma. *N. Engl. J. Med.* **2021**, *385*, 683–694. [CrossRef]

24. Ravaud, A.; Motzer, R.J.; Pandha, H.S.; George, D.J.; Pantuck, A.J.; Patel, A.; Chang, Y.-H.; Escudier, B.; Donskov, F.; Magheli, A.; et al. Adjuvant Sunitinib in High-Risk Renal-Cell Carcinoma after Nephrectomy. *N. Engl. J. Med.* **2016**, *375*, 2246–2254. [CrossRef]

25. Malone, E.R.; Oliva, M.; Sabatini, P.J.B.; Stockley, T.L.; Siu, L.L. Molecular profiling for precision cancer therapies. *Genome Med.* **2020**, *12*, 8. [CrossRef]

26. Dimitrieva, S.; Schlapbach, R.; Rehrauer, H. Prognostic value of cross-omics screening for kidney clear cell renal cancer survival. *Biol. Direct* **2016**, *11*, 68. [CrossRef]

27. Li, P.; Conley, A.; Zhang, H.; Kim, H.L. Whole-Transcriptome profiling of formalin-fixed, paraffin-embedded renal cell carcinoma by RNA-seq. *BMC Genomics* **2014**, *15*, 1087. [CrossRef]

28. Goeman, J.J. L1 Penalized Estimation in the Cox Proportional Hazards Model. *Biometrical J.* **2010**, *52*, 70–84. [CrossRef]

29. Dai, J.; Lu, Y.; Wang, J.; Yang, L.; Han, Y.; Wang, Y.; Yan, D.; Ruan, Q.; Wang, S. A four-gene signature predicts survival in clear-cell renal-cell carcinoma. *Oncotarget* **2016**, *7*, 82712–82726. [CrossRef]

30. Chen, L.; Luo, Y.; Wang, G.; Qian, K.; Qian, G.; Wu, C.L.; Dan, H.C.; Wang, X.; Xiao, Y. Prognostic value of a gene signature in clear cell renal cell carcinoma. *J. Cell. Physiol.* **2019**, *234*, 10324–10335. [CrossRef]

31. Li, F.; Hu, W.; Zhang, W.; Li, G.; Guo, Y. A 17-Gene Signature Predicted Prognosis in Renal Cell Carcinoma. *Dis. Markers* **2020**, *2020*, 8352809. [CrossRef] [PubMed]

32. Zhang, Z.; Lin, E.; Zhuang, H.; Xie, L.; Feng, X.; Liu, J.; Yu, Y. Construction of a novel gene-based model for prognosis prediction of clear cell renal cell carcinoma. *Cancer Cell Int.* **2020**, *20*, 27. [CrossRef] [PubMed]

33. Uhlen, M.; Zhang, C.; Lee, S.; Sjöstedt, E.; Fagerberg, L.; Bidkhori, G.; Benfeitas, R.; Arif, M.; Liu, Z.; Edfors, F.; et al. A pathology atlas of the human cancer transcriptome. *Science* **2017**, *357*, eaan2507. [CrossRef] [PubMed]

34. Jiang, M.; Wang, H.; Chen, H.; Han, Y. SMARCD3 is a potential prognostic marker and therapeutic target in CAFs. *Aging* **2020**, *12*, 20835–20861. [CrossRef] [PubMed]

35. Wei, R.; Liu, H.; Li, C.; Guan, X.; Zhao, Z.; Ma, C.; Wang, X.; Jiang, Z. Computational identification of 29 colon and rectal cancer-associated signatures and their applications in constructing cancer classification and prognostic models. *Front. Genet.* **2020**, *11*, 740. [CrossRef]

36. Zuo, Y.; Xu, X.; Chen, M.; Qi, L. The oncogenic role of the cerebral endothelial cell adhesion molecule (CERCAM) in bladder cancer cells in vitro and in vivo. *Cancer Med.* **2021**, *10*, 4437–4450. [CrossRef]

37. Roldán, F.L.; Lozano, J.J.; Ingelmo-Torres, M.; Carrasco, R.; Díaz, E.; Ramirez-Backhaus, M.; Rubio, J.; Reig, O.; Alcaraz, A.; Mengual, L.; et al. Clinicopathological and Molecular Prognostic Classifier for Intermediate/High-Risk Clear Cell Renal Cell Carcinoma. *Cancers* **2021**, *13*, 6338. [CrossRef]

38. Lundin, L.; Larsson, H.; Kreuger, J.; Kanda, S.; Lindahl, U.; Salmivirta, M.; Claesson-Welsh, L. Selectively desulfated heparin inhibits fibroblast growth factor-induced mitogenicity and angiogenesis. *J. Biol. Chem.* **2000**, *275*, 24653–24660. [CrossRef]

39. Chen, E.; Stringer, S.E.; Rusch, M.A.; Selleck, S.B.; Ekker, S.C. A unique role for 6-O sulfation modification in zebrafish vascular development. *Dev. Biol.* **2005**, *284*, 364–376. [CrossRef]

40. Hellerbrand, C.; Amann, T.; Schlegel, J.; Wild, P.; Bataille, F.; Spruss, T.; Hartmann, A.; Bosserhoff, A.K. The novel gene MIA2 acts as a tumour suppressor in hepatocellular carcinoma. *Gut* **2008**, *57*, 243–251. [CrossRef]

41. Sasahira, T.; Kirita, T.; Nishiguchi, Y.; Kurihara, M.; Nakashima, C.; Bosserhoff, A.K.; Kuniyasu, H. A comprehensive expression analysis of the MIA gene family in malignancies: MIA gene family members are novel, useful markers of esophageal, lung, and cervical squamous cell carcinoma. *Oncotarget* **2016**, *7*, 31137–31152. [CrossRef]

42. Expression of MIA2 in cancer—Summary—The Human Protein Atlas. Available online: https://www.proteinatlas.org/ENSG0 0000150527-MIA2/pathology (accessed on 12 May 2022).

43. Ma, Q.; Wu, K.; Li, H.; Li, H.; Zhu, Y.; Hu, G.; Hu, L.; Kong, X. ONECUT2 overexpression promotes RAS-driven lung adenocarcinoma progression. *Sci. Rep.* **2019**, *9*, 20021. [CrossRef]

44. Wu, Y.; Jiang, G.; Zhang, N.; Liu, S.; Lin, X.; Perschon, C.; Zheng, S.L.; Ding, Q.; Wang, X.; Na, R.; et al. HOXA9, PCDH17, POU4F2, and ONECUT2 as a Urinary Biomarker Combination for the Detection of Bladder Cancer in Chinese Patients with Hematuria. *Eur. Urol. Focus* **2020**, *6*, 284–291. [CrossRef] [PubMed]

45. Lu, T.; Wu, B.; Yu, Y.; Zhu, W.; Zhang, S.; Zhang, Y.; Guo, J.; Deng, N. Blockade of ONECUT2 expression in ovarian cancer inhibited tumor cell proliferation, migration, invasion and angiogenesis. *Cancer Sci.* **2018**, *109*, 2221. [CrossRef] [PubMed]

46. Guo, H.; Ci, X.; Ahmed, M.; Hua, J.T.; Soares, F.; Lin, D.; Puca, L.; Vosoughi, A.; Xue, H.; Li, E.; et al. ONECUT2 is a driver of neuroendocrine prostate cancer. *Nat. Commun.* **2019**, *10*, 278. [CrossRef] [PubMed]

47. Wang, L.; Hu, F.; Shen, S.; Xiao, H.; Li, G.; Wang, M.S.; Mei, J. Knockdown of SOX12 expression inhibits the proliferation and metastasis of lung cancer cells. *Am. J. Transl. Res.* **2017**, *9*, 4003.

48. Gu, W.; Wang, B.; Wan, F.; Wu, J.; Lu, X.; Wang, H.; Zhu, Y.; Zhang, H.; Shi, G.; Dai, B.; et al. SOX2 and SOX12 are predictive of prognosis in patients with clear cell renal cell carcinoma. *Oncol. Lett.* **2018**, *15*, 4564–4570. [CrossRef]
49. Huang, W.; Chen, Z.; Shang, X.; Tian, D.; Wang, D.; Wu, K.; Fan, D.; Xia, L. Sox12, a direct target of FoxQ1, promotes hepatocellular carcinoma metastasis through up-regulating Twist1 and FGFBP1. *Hepatology* **2015**, *61*, 1920–1933. [CrossRef]
50. Gao, Y.; Li, L.; Hou, L.; Niu, B.; Ru, X.; Zhang, D. SOX12 promotes the growth of multiple myeloma cells by enhancing Wnt/β-catenin signaling. *Exp. Cell Res.* **2020**, *388*, 111814. [CrossRef]
51. Ghatalia, P.; Gordetsky, J.; Kuo, F.; Dulaimi, E.; Cai, K.Q.; Devarajan, K.; Bae, S.; Naik, G.; Chan, T.A.; Uzzo, R.; et al. Prognostic impact of immune gene expression signature and tumor infiltrating immune cells in localized clear cell renal cell carcinoma. *J. Immunother. Cancer* **2019**, *7*, 139. [CrossRef]